

25X1A

25X1A

Approved For Release 2004/01/15 : CIA-RDP79M00097A000200010005-5
May 1

SECRET

FIRST THOUGHTS ON AN ALL-SOURCE ELECTRONIC
BIOGRAPHIC INTELLIGENCE CENTER

25X1A

I. Impetus for the Concept of an All-Source Electronic Biographic Intelligence Center

There are numerous files of biographic information maintained by various member agencies of the Intelligence Community and the Federal Government. Each of these files is tailored to the mission and responsibilities of the particular agency maintaining the file. Exploitation of these biographic files falls into two major categories: (1) official U.S. Government name checks, and (2) Intelligence Community biographic research.

The first of these uses requires rapid and thorough searching of all pertinent biographic files in the Federal Government complex.. At present this [] is seldom done thoroughly and rarely done rapidly. It is characteristic of a large proportion of these [] result in a negative report, that is, no information. Some way of expediting this sort of routinized reply to a [] request by automatic or semi-automatic means would permit a tremendous saving in human resources as well as provide an improved name check service.

The second of these uses, i.e., biographic research on personalities of intelligence interest, is substantively much more complex [] In this case the requester is looking for all fragments of available information (both classified and unclassified) about an individual or individuals. These bits and fragments are subjected to analysis from which process an intelligence

SECRET

Approved For Release 2004/01/15 : CIA-RDP79M00097A000200010005-5

SECRET

Approved For Release 2004/01/15 : CIA-RDP79M00097A000200010005²⁵
product or judgment (decision) ensues.

The division of responsibility for biographic intelligence is formalized in various Director of Central Intelligence Directives (DCIDs). "A study of the pertinent biographic directives suggests that two basic policy decisions dictated administrative decentralization in this area. One was a decision to separate biographic files, based on 'collateral' data, by the general occupation of the individuals described therein. The other was to separate personality files of a 'non-collateral' nature by the type of source from which the information was derived. As a consequence of the former, five different agencies are currently engaged in compiling, maintaining and servicing from biographic information files. They are as follows:

25X1A

25X1

The division of substantive responsibility for maintaining each of these files is in some cases artificial and/or overlapping. Duplication of effort or gaps in coverage result. In addition, proliferation of biographic reference points makes the consumers task

1. The Biographic Register Information System: System Description.
Central Intelligence Agency, Office of Central Reference, December 1960.

SECRET

SECRET

Approved For Release 2004/01/15 : CIA-RDP79M00097A000200010005-5
extremely difficult. Not only should he be aware of all the various files he can exploit, he also needs to know whether these files are redundant or complementary in order to place his requests most expeditiously.

Most agencies have frequent need to search the files of at least one other agency maintaining biographic information. Frequently such inter-agency name searches result in a negative report (no information). But in order to deliver the negative report, nearly as much file searching is required as in servicing a positive report. Human resources are sadly depleted by this constant and significantly large demand on human file searching capabilities.

The advantages of a central reference point for requests involving biographic information have been recognized by consumers of this kind of information for a long time. Until recently the attendant problems involved in processing such a large volume of diverse information encompassing all levels of security classification made the task virtually impossible. However, with the advent of new electronic data processing equipment permitting comprehensive indexing in depth and rapid retrieval of large volumes of information, the concept of an all-source electronic biographic intelligence center begins to fall in the realm of visionary planning.

This paper addresses itself to several dimensions of this concept that would need to be considered by systems planners and designers of such a center. The first of these considerations concerns a survey of extant biographic files and a decision as to

25X1A

SECRET

SECRET

Approved For Release 2004/01/15 : CIA-RDP79M00097A000200010005-5
which of these should be included in an all-source system.

(top is)
converted

A second consideration deals with operational and substantive advantages and disadvantages of the concept of an all-source biographic center.

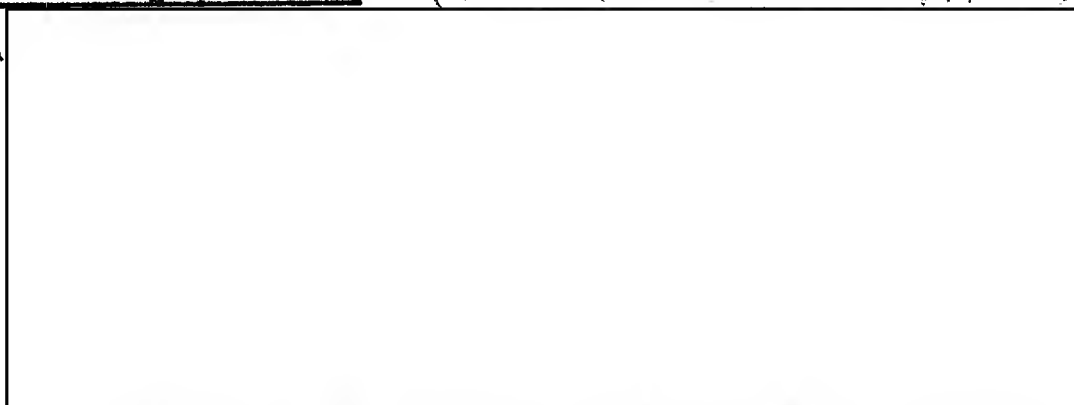
The third consideration addresses itself to consumer requirements for use of such a system. Development of these requirements establishes the guidelines for design of indexing schemes, file organization, types of retrieval capability, interface constraints, etc.

A fourth consideration outlines the problem areas that will inevitably arise in the design of such a system and suggests possible ways for coping with them.

II. Extant Biographic Files (manual or machine supported)

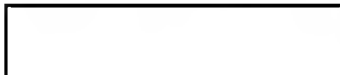
A

25X1



Assigned primary responsibility for biographic information on scientific and technical personalities, and significant economic personalities not covered by the Department of State.²

25X1



SECRET

~~SECRET~~

25X1

STATSPEC

Maintains biographic information files on Agitprop officials, party theoreticians, academicians, and Soviet commentators (includes scattered information on some East European satellite and Chinese Communist propagandists); maintains a catalog of Soviet radio commentators with each one's specialty noted; maintains a card file listing radio lectures by candidate academicians in the social sciences; maintains a card file of authors of articles that have appeared in KOMMUNIST, top theoretical journal of the CPSU.

ILLEGIB

5. GR/OCR/DD/I

Assigned responsibility for maintaining photographic personality files on people of intelligence significance.

6. OS/DD/8

Assigned responsibility for maintaining biographic files on all persons for whom a security investigation has been initiated by CIA.

B. State (Biographic Information Division, Office of Functional and Biographic Intelligence, Bureau of Intelligence and Research)

Assigned primary responsibility for political, social and cultural personalities, and economic personalities of political significance.

C. Army/Navy/Air

Assigned responsibility for maintaining a biographic intelligence file system on foreign military personalities of their counterpart services.

~~SECRET~~

SECRET

D. FBI

Assigned responsibility for maintaining internal security biographic information files on all U.S. citizens, foreign visitors to the U.S., and all persons of known or suspect hostile intent to the U.S.



F. AEC (should it be included?)

G. Naturalization and Immigration Service (should it be included?)

III. All-Source Biographic Center - Advantages versus Disadvantages

A. Advantages

1. The most obvious advantage of an all-source biographic center would be a single, central reference point for all consumers of biographic information in the Federal Government and the Intelligence Community. A requester would place his request only once; the response to his request would include all information available to him throughout the Federal Government and/or the Intelligence Community.
2. A single biographic center would eliminate or at least reduce overlapping efforts in several agencies to index and control the same biographic information. It would also point up gaps in biographic coverage.
3. Time consumed in requesting documents and/or enclosures from other agencies would be eliminated.

*Possible with
an all source
file.*

SECRET

25X1

Approved For Release 2004/01/15 : CIA-RDP79M00097A000200010005-5

Approved For Release 2004/01/15 : CIA-RDP79M00097A000200010005-5

SECRET

B. Disadvantages

1. The centralization of all biographic files in one location would preclude the maintenance of biographic information repositories in any of the agencies participating in the central system. Any duplication of extant files would be both inefficient and uneconomical. The importance of this point is that centralization would deprive each participating agency of autonomous control over its biographic reference function for internal support. They would be forced to depend upon the central repository. Unless the central facility really provided timely and comprehensive reference service, the participating agency would find itself with no alternative capability to retrieve biographic information. (One way to safeguard against this possibility might be to only centralize the indexes to extant biographic files. The actual repositories of biographic information would remain with the participating agencies who would then have the capability to utilize their own files for internal support as well as profit from the availability of a central biographic reference service for external requests.)
2. In order for a central biographic reference system to be effective, the index to file holdings (whether these files are centrally located or decentralized) would need to be both comprehensive in fields of information controlled and

*agencies
holding
files in
their own
reference*

SECRET

SECRET

omniscient in depth of indexing employed. The quality of reference service provided by a centralized system would be dependent upon the speed and accuracy with which participating agencies supplied new, revised, or purged index entries to the central system. If the central system could not provide satisfactory reference support, each agency would rely on its own biographic holdings, vitiating any potential advantages a central reference facility could provide. Member agencies would then have returned to the old decentralized system but with the appendage of an expensive, emasculated central facility ignored by the consumers it was designed to serve.

ILLEGIB

3. Unless operationally feasible and foolproof methods for matching the security classification of documents to the security clearance of the requester can be designed, effective central reference service would hang up on security screening delays before release of materials to requesters.
4. In order for an electronic central reference facility to provide effective service, a uniform machine-readable language must be developed for indexing and retrieval. A requester unfamiliar with this language and remotely located from this central reference point may have difficulty in selecting the best way to phrase his request. This communication barrier between the requester and the

SECRET

and the central biographic index may require the assistance of intermediaries who can translate the requester's question into appropriate terms for precise machine searching. Biographic specialists or reference technicians may be necessary to bridge this gap.

5. Centralizing and standardizing indexing control over biographic holdings may work a hardship on certain agencies who ordinarily would only superficially index their biographic material. Forcing them to conform to a standard indexing in depth procedure would create an extra workload they may not be willing to assume. The alternative is to have the indexing function done at the central facility.

Time consuming.
Tremendous increase in personnel in either instance

SECRET

SECRET

IV. Designing the System

A. Determining the Limits of the Data Base

The first step in designing an automated central biographic reference system requires a determination of 1) all extant biographic files, 2) the size and nature of the indexes to these files, and 3) characteristics of the document holdings themselves (e.g., dossiers, hard copy bulletins and reports, aperture cards, microfilm, positive and negative photos or film strips etc.).

B. Determining Consumer Requirements

The next step requires an investigation of how consumers would optimally use these files. All classes of requests must be specified and the volume of requests in each class determined.

C. Basic Factors in Any Storage and Retrieval System

Any information storage and retrieval system must consider several problem areas in the design phase: 1) the information control problem, 2) the physical storage problem, and 3) the retrieval problem. In the information control area, several basic questions must be answered:

1. Is the system to be an information storage and retrieval system or a document storage and retrieval system or both?
2. What categories of information are important enough to control for every document indexed?
3. What type of indexing scheme(s) is/are to be employed?
(hierarchical codes, classificatory or taxonomic codes,

SECRET

coordinate indexing, uniterms, key words, a combination of these, or some new technique)

4. How much modifying information is to be included in the index? (linkage analysis, modifiers, descriptors, action and use codes, directional indicators, tags, cross references, see also entries, etc.)

5. If alphanumeric codes are used, should the system be able to convert them to their clear text equivalents?

6. How does the system provide for consistent indexing of the same kinds of information? (inter and intra-coder reliability)

7. How much of the indexing can be done automatically, i.e., by computer dictionary look-up procedures?

8. How is name verification to be accomplished?

9. When does the name verification process occur? At time of input or at time of servicing a request?

10. Can the input format be controlled at the source originating the document?

11. If input is automatic, can the input processing equipment perform "legality checks" on errors in format?

The physical storage problem requires answers to these basic questions:

1. How should the document holdings be stored? (hard copy, aperture cards, document images, microfilm, magnetic tape, punched paper tape, punched cards, random access storage, etc.)

SECRET

SECRET

2. If the index is to be mechanized, in what form will it be maintained? (punched cards, magnetic tape, random access storage, core memory, photostatic disc, etc.)
3. In what format(s) is the indexed information to be stored?
4. In what sequence(s) is the indexed information to be stored? ✓
5. Should all index entries be stored in one file or partitioned among several files? ? *Explanations*
6. What kinds of physical security are required to protect the data base and the index from getting into the hands of un-cleared personnel?
7. What kind of file backup is required to protect against loss of the data base and the index by catastrophe?
8. Is the data base and index stored in a form (machine readable or otherwise) that is compatible with users' equipment and/or needs? (the interface problem)
9. Is it desirable to purge the file of old or obsolete information at periodic intervals?

check this

check this

check this

ADD this

The retrieval problem requires answers to these questions:

1. Does the index storage sequence(s) permit rapid retrieval of the major information categories?
2. Can the index entries selected by a search be listed in clear text and in a variety of sequences for easier use by the requester?
3. Once a search of the index is made, are the documents referenced easy to locate?

SECRET

SECRET

4. Does the index system lend itself to data extraction and refinement?
5. Can a document or information be recovered from the file without removing it from the file thus making it unavailable for other requests?
6. Is it desirable to incorporate into the index evaluative comments from previous requesters about quality of the information or its interpretation so that a new requester can take advantage of previous analysis in his area of interest? (evaluative feedback to the system)

D. File Organization

1. Systems of File Organization

- a. The Unit Record File. All the information indexed from one document constitutes a single index entry or unit record. Depending upon the physical medium employed to store the index entry, this unit record may be bound by equipment limitations. Punched cards are limited to 80 columns of information which must be allocated on the card to fixed length fields for ease of search, selection, and sequencing by EM equipment. Data processing computers permit variable length unit records with both fixed and variable length fields.
- b. The Unit Record File with Internal Coordinate Indexing. All the information indexed from one document constitutes

SECRET

SECRET

-15-

a single index entry or unit record. The index is stored in sequence by a primary field. A second index is constructed on a secondary field indicating which entries in the primary sequence contain each kind of information displayed in the secondary field. An example of this kind of file organization might be a file of biographic information about all scientists attending international meetings. The primary sequence of this file would be maintained by name. A second index could be constructed relating to a secondary field, say, institution employed by. The names of all institutions occurring in the primary index would appear in the secondary index in alphabetic order. Along with each institution in the secondary index all entries in the first index mentioning that institution would be cross referenced.

25X1

25X1A

- c. The Multi-segment File with Unit Record Input and Internal Cross Referencing. All the information indexed from one document constitutes a single unit record for input. However, the format of the unit record is

SECRET

SECRET

partitioned into segments, each controlling different kinds of information. As the unit record is read through the input processing equipment, it is fractionized into these segments and stored in separate internal files. The processing equipment provides internal cross referencing addresses so that the original unit record can be reassembled or segments can be related. This type of file organization facilitates retrieval when unit records are long and variable and when most requests involve searching only part of the fields in a record.

- d. The Multi-level File. This file structure is organized into several levels like a pyramid. The first or top level usually contains only timely or critically important information and is stored in high speed random access memory. The second level contains additional information elaborating on what is contained in the first level file and is usually stored in random access memory or on magnetic tape. The third level contains further pertinent but less essential information such as historical background, related documents, previous patterns, evaluations by other consumers, summaries, etc. The three file levels can be tied together by cross referencing so that a requester can move down through the file structure in as much

SECRET

SECRET

-17-

depth as he requires. This kind of file organization is especially suited to man-machine systems where the analyst interrogates the file, receives an answer, generates a new question, re-interrogates the file, etc. until he has satisfied all leads. The third level file is most economically stored in document image form such as on Minicards, aperture cards, film strips, or micro-film.

2. Factors to be Considered in Determining File Structure and Organization

- a. Scope of the information to be controlled (breadth of the data base)
- b. Volume of information to be controlled
- c. Rate of retirement of purged information from the file
- d. Depth of indexing required
- e. Maximum physical length of a unit record
- f. Average physical length of a unit record
- g. Variety of input formats required
- h. Access time to information stored in different kinds of memory (core, RAMAC, magnetic tape, magnetic drum, document image, etc.)
- i. Sophistication and speed of search logic
- j. Relative costs

3. Desirable Features in File Structure and Organization

- a. Simplicity and ease of input (automatic input processing

SECRET

SECRET

Approved For Release 2004/01/15 : CIA-RDP79M00097A000200010005-5

being the ultimate goal)

- b. Simplicity and accuracy of internal editing and filing
- c. Search flexibility
- d. Speed of search
- e. Quality of search (few "false drops", no misses)
- f. Adaptability to change (dynamic and fluid file design)
- g. Standardized indexing and request language
- h. Incorporation into the file of evaluative feedback from users of the system.

E. An All-Source Electronic Biographic Intelligence Center: Some Tentative Charts

Figure 1 is a graphic representation of the concept of a central Biographic Intelligence Center serving the needs of member agencies of the Intelligence Community and the Federal Government.

Figure 2 concerns itself with the relationships between the logical components of an all-source electronic biographic information system. This figure is designed to suggest possible kinds of input to and output from a central biographic index system. The actual document files may be decentralized.

Index entries prepared by participating agencies may enter the system by direct communication link from the field station reporting or by punched cards prepared in the Washington area or elsewhere. A stored computer program would process the input into the various index files, performing necessary

SECRET

Approved For Release 2004/01/15 : CIA-RDP79M00097A000200010005-5

~~SECRET~~

-19-

information control operations such as checking the legality of formats, augmenting the indexes by addition of information from dictionaries and gazetteers, and providing internal cross referencing for retrieval purposes.

Reference service from the central system could take the form of lists of document references and where they can be located, listings of selected index entries arranged in various sequences, selected information on magnetic tape or punched cards for use in another data processing system, or paper tape for transmitting name check responses to the field via direct communication link.

~~SECRET~~

25X1

Approved For Release 2004/01/15 : CIA-RDP79M00097A000200010005-5

Approved For Release 2004/01/15 : CIA-RDP79M00097A000200010005-5

SECRET

-21-

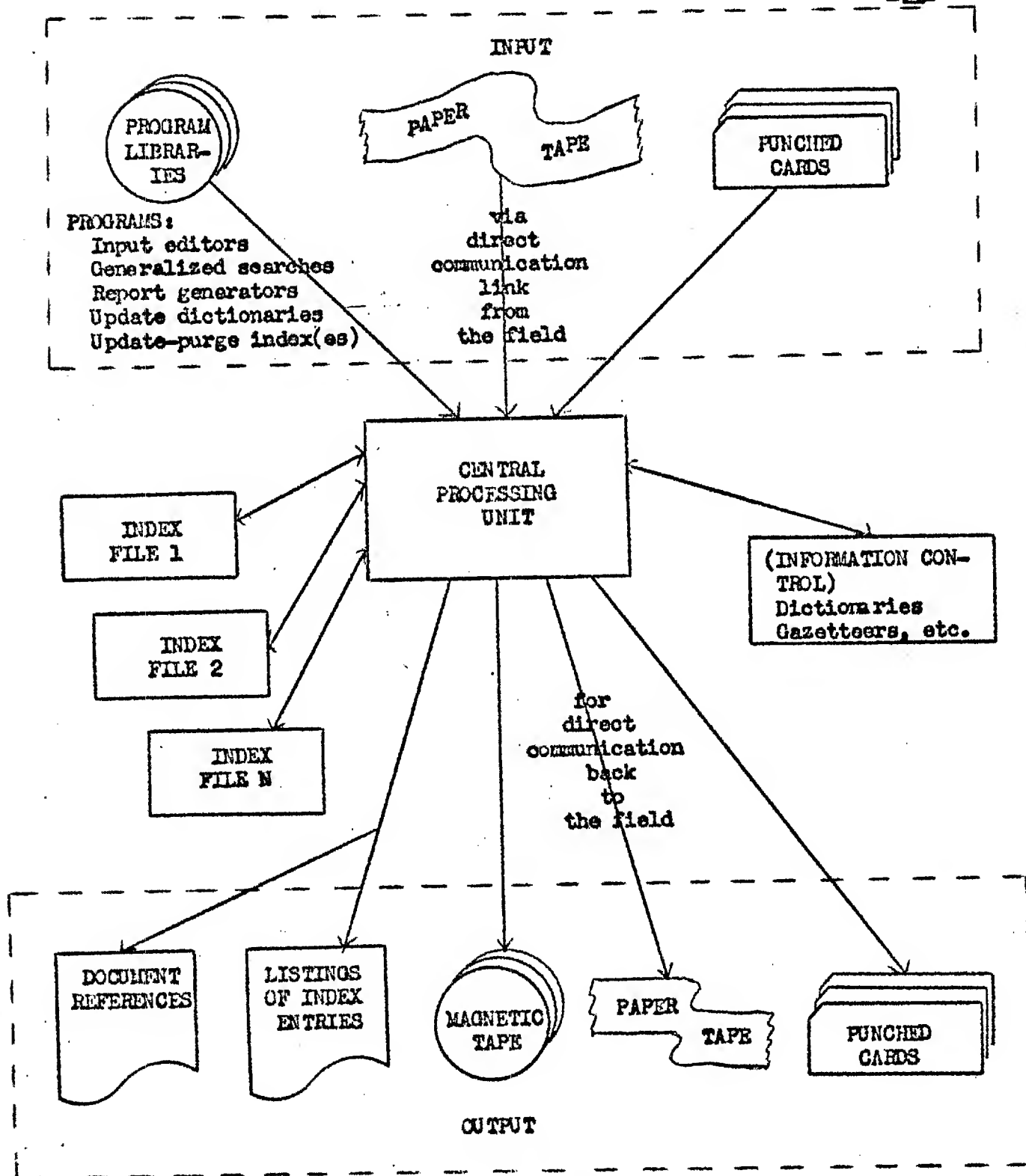


Figure 2

Logical Components of An Electronic
Biographic Intelligence Information System

SECRET

V. All-Source Electronic Biographic Intelligence Center: Problem Areas

A. Security Safeguards

1. Physical Site Security. The computer site should be established with a security environment corresponding to the highest level of security classification included in the document holdings. All operating and maintenance personnel should have this level of clearance. This arrangement would avoid the daily security problems connected with making sure that all input-output devices, core memory, ancillary storage, buffer units, and console had been cleared of classified information after each scheduled operation.
2. Communications Security. The configuration of on-line and off-line equipment at the computer site plus any direct communication links with satellite users would have to be made secure against compromising emanations from line or air transmissions. This would include any Flexowriter-type equipment used in the field to prepare input to the central system.
3. Top Secret Control. Any information entering the system with unknown security classification should be treated as though it had the highest classification until otherwise confirmed.

The problem of how to treat a collection of top secret documents all contained on one reel of magnetic or paper tape must be resolved. The clerical burden of logging in and out all of these individual documents everytime a reel of tape is moved

SECRET

~~SECRET~~

-23-

from one person's jurisdiction to another's must be avoided and some other less cumbersome form of security control provided.

4. Personnel Security.

- a. Personnel Operating the Center. All operating, maintenance, and necessary support personnel assigned to the Center should have the highest level of security classification used in the Center.
- b. Consumers. A foolproof system for matching the security level of information provided to a consumer with the consumer's security clearance must be assured by automatic means. This certainly involves including a field for security classification in the format(s) for each index entry so that the computer programs used in searching the index will select only that information which the requester is entitled to see.

If there is additional information in the system of a higher security classification than the requester is entitled to see, a supervisory person with the higher security clearance could be informed of this fact. He could then decide if the person making the request was being deprived of vital information he needed to know and, consequently, this supervisory person could take steps to initiate the higher level security clearance needed by the requester.

~~SECRET~~

B. Variant Input Languages

When the origination of biographic information cannot be controlled at the source, it may be received by the system in a variety of languages. Some solution to this problem must be found. It probably is not feasible to translate all foreign language document receipts into English before processing them into the system. Perhaps a better solution would be to prepare the index entry in English but leave the document or source referenced as it was received. This leaves the burden of translation on the requester before he can exploit the information contained in the document or source. It also would require indexers with multi-language capability.

C. Variant Name Spellings

The problem of when and how to verify the identity of a person named in a document must be resolved. In some systems every name is verified before it enters the system. This is required so that the document or source can be filed in the dossier or section of a card file along with other references to the same individual. At the time of servicing a request for this individual, retrieval is a fairly simple task. Once the proper dossier number is determined by a biographic locator system, all references to the individual are already physically assembled in one place.

SECRET

~~SECRET~~

-25-

Systems that incorporate name verification at the input phase may create unnecessary work. Much of the information in a file may lie dormant and never be of interest to requesters. In time, this material would be retired from the active file because of date of information, low circulation rate, or some other criterion. The time spent in verifying who the individual is has been wasted.

not there

25X1A

Other systems reserve name verification to the output phase (e.g.,). The name in a document or source is not verified until it is of interest to some requester. Thus, documents which have no interest among consumers get the minimum amount of input processing required to get them into the system. This appears to be a more economical way to handle the variant name problem providing that the verification process at the output stage is not so time-consuming that poor reference service results. Time-consuming name verification at output can be alleviated by storing index entries for biographic references with similar names in common groups. Establishing these name groups requires intensive study and understanding of transliteration of names from one language to another, common misspellings or corruptions of names, alternative spellings, etc. But once inclusive name groups have been established, servicing a request involves only the examination of a manageable group of biographic index references rather than the entire file.

~~SECRET~~

D. Debilitation of the Central System

Unless effective administrative controls are established to assure timely and efficient support to the Central System by participating agencies, it will atrophy from lack of comprehensive input and, consequently, lack of use.

SECRET